

The Swift satellite and redshifts of long gamma-ray bursts

Z. Bagoly¹, A. Mészáros², L.G. Balázs³, I. Horváth⁴, S. Klose⁵, S. Larsson⁶, P. Mészáros⁷, F. Ryde⁶, and G. Tuszáný⁸

¹ Laboratory for Information Technology, Eötvös University, H-1117 Budapest, Pázmány P. s. 1/A, Hungary
e-mail: zsolt.bagoly@elte.hu

² Astronomical Institute of the Charles University, V Holešovičkách 2, CZ 180 00 Prague 8, Czech Republic
e-mail: meszaros@mbbox.cesnet.cz

³ Konkoly Observatory, H-1525 Budapest, POB 67, Hungary
e-mail: balazs@konkoly.hu

⁴ Department of Physics, Bolyai Military University, H-1456 Budapest, POB 12, Hungary
e-mail: horvath.istvan@zmne.hu

⁵ Tautenburg Observatory, D-07778 Tautenburg, Sternwarte 5, Germany
e-mail: klose@tls-tautenburg.de

⁶ Stockholm Observatory, AlbaNova, SE-106 91 Stockholm, Sweden
e-mail: stefan@astro.su.se, felix@astro.su.se

⁷ Dept. of Astronomy & Astrophysics, Pennsylvania State University, 525 Davey Lab. University Park, PA 16802, USA
e-mail: nnp@astro.psu.edu

⁸ Rényi Institute of Mathematics, Hungarian Academy of Sciences, H-1364 Budapest, POB 127, Hungary
e-mail: tusnady@renyi.hu

Received 7 October 2005; accepted

Abstract. Until 6 October 2005 sixteen redshifts have been measured of long gamma-ray bursts discovered by the Swift satellite. Further 45 redshifts have been measured of the long gamma-ray bursts discovered by other satellites. Here we perform five statistical tests comparing the redshift distributions of these two samples assuming - as the null hypothesis - identical distribution for the two samples. Three tests (Student's t -test, Mann-Whitney test, Kolmogorov-Smirnov test) reject the null hypothesis on the significance levels between 97.19 and 98.55%. Two different comparisons of the medians show extreme (99.78 – 99.99994)% significance levels of rejection. This means that the redshifts of the Swift sample and the redshifts of the non-Swift sample are distributed differently - in the Swift sample the redshifts are on average larger. This statistical result suggests that the long GRBs should on average be at the higher redshifts of the Swift sample.

Key words. gamma-rays: bursts – Cosmology: miscellaneous

1. Introduction

Recently, the Swift satellite (Gehrels et al. 2005) detected the gamma-ray burst GRB050904, for which the redshift is $z = 6.29$, directly measured from the afterglow data (Haislip et al. 2005a, Haislip et al. 2005b, Antonelli et al. 2005, Kawai et al. 2005). In addition, it is remarkable that the GRBs detected by the Swift satellite seem to have systematically bigger redshifts on average than the redshifts of GRBs detected by other satellites (Greiner 2005, Friedman & Bloom 2005). Hence, the question emerges immediately: does the redshift distribution of GRBs detected by Swift significantly differ from the other GRBs with known redshifts?

The search for the answer to this question is the first aim of this paper: using statistical analyses we show that this is really the case. We briefly discuss the importance, meaning and the consequence of this statistical result on the redshift distribution of long GRBs. The detailed discussion of the astrophysical reasons for the obtained results is the topic of a forthcoming paper.

2. The samples

Greiner (2005) lists the observations concerning the afterglows of GRBs, and among them he also selects and lists the confirmed redshifts (see the reference therein and also Friedman & Bloom 2005). For our statistical studies this survey and selection was used. There are 16 bursts

Send offprint requests to: A. Bagoly

with measured redshifts, which were detected first by the Swift satellite, during the period 1 January - 6 October 2005. These 16 redshifts define here the "Swift" sample, where the smallest redshift is $z = 0.26$ (GRB050714, Prochaska et al. 2005); the largest redshift is $z = 6.29$ (GRB050904, Kawai et al. 2005); the mean redshift is $\bar{z}_{Swift} = 2.42$; the variance is $\sigma_{Swift}^2 = 2.60$; and the median is $z_{Swift,med} = 2.405$.

There are a further 45 redshifts in the Greiner's table from the period 28 February 1997 - 6 October 2005. In the cases GRB011030X and GRB980329 there are upper limits only, and these cases were not included into the sample; in three cases (GRB020305; GRB991216; GRB980326) the redshifts were only estimated, but here we considered them; for GRB000214 the estimated redshift is between 0.37 and 0.47, and here $z = 0.42$ was used (Antonelli et al. 2000). In this way we obtain the "non-Swift" sample containing 45 redshifts. In this sample the smallest redshift is $z = 0.0085$ (GRB980425, Galama et al. 1998); the largest redshift is $z = 4.50$ (GRB000131, Andersen et al. 2000); the mean redshift is $\bar{z}_{non-Swift} = 1.31$; the variance is $\sigma_{non-Swift}^2 = 1.11$; and the median is $z_{non-Swift,med} = 1.02$.

It should be noted that both samples probably only constitute long GRBs ($> 2s$), which should be different from short and intermediate bursts (Horváth 1998, Meegan et al. 2001, Horváth 2002, Balázs et al. 2003, Hakkila et al. 2003, Borgonovo 2004). In other words, this paper deals exclusively with long GRBs.

3. The statistical tests and their results

The two samples were compared using the parametric Student's t -test and non-parametric rank based statistical tests: the Mann-Whitney test, Kolmogorov-Smirnov test and two median tests. The rank based tests has the clear advantage of being unaffected by any monotonic transformation in the z values.

The most common test for the comparison of the average redshifts is the Student's t -test. The details of this test together with the calculation of the significance level can be found in Chap. 14.2 of Press et al. (1992). Note here that, since the variances are different, one has to use the modification of the test for unequal variances. The relevant formulas are given by Eq.(14.2.3) and Eq.(14.2.4). One obtains $t = 2.57$ for the degree-of-freedom $dof = 19.75$. The Student's t -test's significance of rejecting the null hypothesis (i.e. the two samples have identical mean values) is 98.15%.

The Mann-Whitney test uses, instead of the values of the redshifts, their ranks. Its meaning with the relevant formulas is explained, for instance, by Lowry (1999). Not going into the details we remark only that the key idea of the test is the following: one considers first the two samples together, and in this common sample, containing in our case 61 objects, one sorts the redshifts into a mono-

tonic increasing sequence. Having this, not the values of redshifts, but only the ranks will be used for both in the non-Swift and in the Swift sample. This means that the null hypothesis is the assumption that in both samples the mean ranks are identical, and the difference comes from chance.

In our case the Mann-Whitney test gives a 98.55% significance level, i.e. it rejects the null hypothesis almost at the same significance level as that of the Student's t -test.

The Kolmogorov-Smirnov test compares the cumulative distributions of the redshifts in the two samples. The details of the test are described in Chap. 14.3 of Press et al. (1992). It gives the significance 97.19%, which means that this test rejects the null hypothesis at this level.

Not only the mean redshifts in the two samples, but also the medians could be different. To verify this expectation we provided the following two tests.

We considered the two samples with all together 61 redshifts, sorted in ascending series. There are 6 non-Swift and 8 Swift GRBs with redshifts larger than $z_{Swift,med} = 2.405$; there are then 39 non-Swift and 8 Swift GRBs with redshifts smaller than $z_{Swift,med}$. In other words, the median of Swift sample separates the non-Swift in a ratio 6 : 39, instead of the expected 1 : 1.

Assuming a binomial distribution with $N = 45$ and $p = 0.5$ the expected most probable value is $Np = 22.5$ (e.g. Mészáros 1997, Balázs et al. 1998) and the expected dispersion is $\sigma_{theor.}^2 = Np(1-p) = 11.25$; i.e. $\sigma_{theor.} = 3.4$. The difference between the most expected and the real value is $39 - 22.5 = 16.5 = 4.9\sigma_{theor.}$. Because both Np and $\sigma_{theor.}$ are much greater than one, the Gaussian approximation of the binomial distribution is allowed. This means that the null hypothesis (i.e. the two medians are the same) is rejected on the 4.9 sigma level, corresponding to the significance 99.99994%.

A modification of this test compares the medians of the Swift and non-Swift samples as follows. Be chosen 16 objects randomly from the sample of $N = 61$ events. In this subsample the mean between the 8th and 9th objects gives the median. Let this 8th (9th) object from the subsample be in the q^{th} (s^{th}) position in the ordered sample of N events ($q < s$). Obviously, it can be $8 \leq q \leq 53$, and $9 \leq s \leq 54$. Then the corresponding probability for the given q and s is simply the number of the good configurations divided by the number of all configurations:

$$P(q, s) = \binom{q-1}{7} \binom{N-s}{7} / \binom{N}{16}, \text{ where for non-negative}$$

integer x and $y \geq x$ we have $\binom{y}{x} = y!/(x!(y-x)!)$. For the median of the Swift sample the corresponding significance level is $\sum_{(q+s) < 95} P(q, s) = 99.78\%$. (The sum must be taken for all combinations with $(q+s)/2 < 47.5$, because for these combinations the medians of the chosen subsamples are smaller than the Swift's median.) The null hypothesis is again *rejected* on this level. We verified this high significance level also by 100000 Monte-Carlo simulations, and confirmed the result above.

4. Discussion and conclusions

All five statistical tests reject the null hypothesis that the redshift distributions for the Swift and non-Swift samples are identical on the significance higher than 95%. The redshifts of the Swift sample are on average larger than that of the non-Swift sample.

Three questions emerge immediately from this result: I. Is this statistical result an important new result indeed? II. What is the reason for this behavior? III. And what is the impact of this result on the redshift distribution of the long GRBs?

First of all, it may seem that the result is in essence not unexpected and is "reasonable". This point of view may follow from the fact that the Swift satellite is surely more sensitive, and it allows for a faster detection of fainter bursts than that of the other satellites. Also the number of detected GRBs during a given time interval is larger. Hence, because "Swift detects more and fainter GRBs, it is reasonable that it also detects higher redshifts on average" - one may think.

Concerning this point of view two important remarks are needed.

First, it is a triviality that either an exact confirmation or an exact rejection of a theoretical expectation - done purely from the observational data by strict statistical methods - is useful. This was done here. In fact, it is not so sure immediately without a detailed statistical testing that the redshifts of the Swift sample are in average bigger than that of the non-Swift sample. For example, if one takes *only* six typical redshift values (minimal, mean and maximal redshifts from both samples) and one does *not* take also the variances, then one simply cannot obtain from these six redshifts *alone* the conclusion that the two distributions have different redshifts in average. The six typical redshifts *alone* are simply not enough for such a claim. It is not even clear what is the "most natural" parameter of the distance: z , $1 + z$, $\log(1 + z)$ or even the luminosity distance? Either could be a "good" parameter: however they will show totally different distribution (mean, variance, skewness, etc.), and it is hard to use parametric tests only. The rank based tests are clearly preferred as they have the clear advantage of being free from such problems. The relevance of our result for the redshift distribution of long GRBs is straightforward: it excludes the possibility that the Swift sample and the non-Swift sample of GRBs originate, on average, in the same redshift range. All this answers the question I.

Second, one should point out that, strictly from the observational point of view, it is simply not yet certain that fainter, long bursts are on average at larger redshifts. The redshifts of long GRBs are still poorly measured. Beyond a few dozens of directly measured redshifts - the ones studied in this paper - only indirect estimations are known for the redshifts of long GRBs. The estimates from the BATSE data using the gamma range alone (Mészáros & Mészáros 1996, Horváth et al. 1996, Reichart & Mészáros 1997,

Lamb & Reichart 2000, Schmidt 2001, Schaefer et al. 2001, Lloyd-Ronning et al. 2002, Norris 2002, Bagoly et al. 2003, Atteia 2003, Lin et al. 2004) suggest that large redshifts (up to $z = 20$ (Mészáros & Mészáros 1996)) for fainter GRBs may well be expected indeed. Also the analyses from the afterglow data suggest (Berger et al. 2005) that the fainter, long GRBs really should on average be at higher redshifts. The very high GRBs' redshifts are discussed in detail by Lamb & Reichart (2000); the cosmological aspects of GRBs - fainter ones expected to be at higher redshifts - are studied, e.g., by Friedman & Bloom (2005). Hence, the theoretical expectation "fainter = larger redshifts on average" is reasonable indeed. But, strictly, there has been no observational certainty yet, and the observational verification of the theoretical expectation has been needed. In essence, just this has been provided in this article using the exact statistical arguments. All this means that the question II. may be answered as follows: The reason for this behavior *can be* instrumental. The instruments on the Swift satellite are more sensitive than the other GRB detectors, and the fainter GRBs should indeed be on average at higher redshifts.

Hence, we allow to conjecture already now that the difference in the average redshifts originates - either partly or fully - from instrumental selection effects. For a more concrete answer a detailed study of the instrumental properties of the considered satellites is needed, which is planned to be provided in a forthcoming paper. It should quantify the importance of the instrumental biases.

Concerning the impact on the redshift distributions, from the results discussed above, three possibilities can be envisaged for the GRB distribution:

A. The long subgroup of GRBs is unique (no further sub-grouping is needed), and its redshift distribution is represented by the non-Swift sample; the redshifts of the Swift sample are biased. We believe that this possibility is strongly disfavored, because the distribution of GRBs before the Swift era were strongly biased (Mészáros et al. 2004), and in the non-Swift sample only the brightest GRBs should be present. Hence, it is hard to imagine that just the non-Swift sample represents the true intrinsic redshift distribution of the unique long GRBs (Mészáros et al. 2004). Also Berger et al. (2005) rejects this possibility.

B. The long subgroup of GRBs is unique, and its redshift distribution is represented more or less well by the Swift sample; the redshifts of the non-Swift sample are biased. It is shown recently (Jakobsson et al. 2005) that the Swift redshift data may represent the real intrinsic redshift distribution of long GRBs, but one cannot exclude the possibility that the exact GRB redshift distribution is neither properly represented by the Swift sample nor by the non-Swift sample. E.g. the redshift distributions of the BATSE sample (Mészáros & Mészáros 1996, Schaefer et al. 2001, Lloyd-Ronning et al. 2002, Bagoly et al. 2003) may be quite different to that of the Swift sample. Lin et al. (2004)

even claims that the majority of GRB in the BATSE sample are at $z > 10$.

One should remark that there are 16 Swift GRBs with measured z and optical transient (OT) from the 32 Swift GRBs with OT observation (50%), while for the non-Swift sample there are 38 GRBs with z and OT from the total of 65 GRBs with OT (58%, Greiner 2005). (Seven non-Swift events have measured z without an OT.) The remarkably higher OT detection rate for the Swift bursts ($\simeq 43\%$ vs. the non-Swift $\simeq 29\%$) indicates that the lower redshift measurement success ratio could be an indication of an observational bias from the cosmic reddening: i.e. for the significantly higher Swift redshifts a spectral line may drop out of the V and R bands, and are detectable only in the J, K bands or at even longer wavelengths. This clearly reduces the success of the redshift determination. The OT search efficiency is clearly $< 100\%$ due to observational constraints, so the $\simeq 43\%$ rate of the OT detection in the Swift era may suggest that the majority of the GRBs are within the optically transparent cosmological region.

C. The long subgroup of GRBs is not necessarily unique, i.e. the different redshifts in our samples may originate from different types of sources with different redshift distributions: e.g. the non-uniqueness of the long subgroup is also an alternative (Borgonovo 2004). Also Berger et al. (2005) considers, as a possibility, that the Swift GRBs with higher redshifts represent a luminous subgroup of the long class alone. On the other hand, recent statistical analyses show (Horváth et al. 2005) that three subgroups of GRBs are enough to explain the statistical properties of the BATSE sample. Hence the long subclass should not be further divided. It means that the observed differences might be accounted only for the different observational strategies of the different experiments.

However, simply the different energy sensitivity of the different spacecrafts could not explain satisfactorily the differences in the redshifts. The Swift and non-Swift trigger energy ranges overlap: in the non-Swift samples' there are 27 "softer" and 6 "harder" GRBs ("softer" and "harder" means that the observed trigger energy is below or above the average Swift trigger energy), while 12 GRBs' trigger energy was within the Swift band. E.g. only $\approx 13\%$ of GRBs were triggered above the Swift energy band, and the median z of this small subsample is higher ($z_{med} = 1.64$) than the non-Swift median $z_{non-Swift, med} = 1.02$. The higher trigger energy may select GRBs of higher redshifts, however, this effect could be even a small sample effect only.

Thus, in our opinion, both the B and C cases are quite possible, and a further detailed study is needed. This we plan to provide in a separate paper. Here we can state only as an answer to question III. that our statistical arguments relying on the observational data alone suggest that the average redshifts of the long GRBs can be great indeed ($z \simeq (2.4 - 2.6)$ or even larger).

Acknowledgements. This study was supported by Hungarian OTKA grant T48870 and NASA grant NAG5-13286. Useful remarks of an anonymous referee are acknowledged.

References

- Antonelli, L.A., Piro, L., Vietri, M., et al. 2000, *ApJ*, 545, L39
 Antonelli, L.A., Grazian, A., D'Avanzo, P., et al. 2005, *GCN Circ.* 3924
 Atteia, J.-L. 2003, *A&A*, 407, L1
 Andersen, M. I., Hjorth, J., Pedersen, H., et al. 2000, *A&A*, 364, L54
 Bagoly, Z., Csabai, I., Mészáros, A., Mészáros, P., Horváth, I., Balázs, L.G. & Vavrek, R. 2003, *A&A*, 398, 919
 Balázs, L.G., Mészáros, A., & Horváth, I. 1998, *A&A*, 339, 1
 Balázs, L.G., Bagoly, Z., Horváth, I., Mészáros, A., & Mészáros, P. 2003, *A&A*, 401, 129
 Berger, E. et al. 2005, *astro-ph/0505107*
 Borgonovo, L. 2004, *A&A*, 418, 487
 Friedman, A.S. & Bloom, J.S. 2005, *ApJ*, 627, 1
 Galama, T. J., Vreeswijk, P. M., Pian, E., et al. 1998, *IAU Circ.*, 6895, 1
 Gehrels, N., Chincarini, G., Giommi, P., et al. 2004, *ApJ*, 611, 1005
 Greiner, J. 2005, <http://www.mpe.mpg.de/~jcg/grbgen.html>
 Haislip, J., Reichart, D., Cypriano, E., et al. 2005a, *GCN Circ.* 3914
 Haislip, J., Nysewander, M., Reichart, D., et al. 2005b, *GCN Circ.* 3919
 Hakkila, J., et al. 2003, *ApJ*, 582, 320
 Horváth, I. 1998, *ApJ*, 508, 757
 Horváth, I. 2002, *A&A*, 392, 791
 Horváth, I., Mészáros, P. & Mészáros, A. 1996, *ApJ*, 470, 56
 Horváth, I. et al. 2005, *A&A*, in press, *astro-ph/0509909*
 Jakobsson, P. et al. 2005, *astro-ph/0509888*
 Kawai, N., Yamada, T., Kosugi, G., et al. 2005, *GCN Circ.* 3937
 Lamb, D.Q., & Reichart, D.E. 2000, *ApJ*, 536, 1
 Lin, J.R., Zhang, S.N. & Li, T.P. 2004, *ApJ*, 605, 819
 Lloyd-Ronning, N.M., Fryer, C.L. & Ramirez-Ruiz, E. 2002, *ApJ*, 574, 554
 Lowry, R. 1999, <http://faculty.vassar.edu/lowry/ch11a.html>
 Meegan, C.A., et al. 2001, *Current BATSE Gamma-Ray Burst Catalog*, <http://gamma-ray.msfc.nasa.gov/batse/grb/catalog>
 Mészáros, A. 1997, *A&A*, 328, 1
 Mészáros, A. & Mészáros, P. 1996, *ApJ*, 466, 29
 Mészáros, A., Bagoly, Z., Balázs, L.G., Horváth, I. & Klose, S. 2004, in *Proc. of the Third Roma Symposium "Gamma-Ray Bursts in the Afterglow Era"*, eds. M. Feroci, F. Frontera, N. Masetti, L. Piro, ASP Publ. Co., p.118
 Norris, J.P. 2002, *ApJ*, 579, 386
 Press, W.H., et al. 1992, *Numerical Recipes, Second Edition*, Cambridge Univ. Press, Cambridge
 Prochaska, J.X., Chen, H.-W., Bloom, J.S., Stephens, A., et al. 2005, *GCN Circ.* 3679
 Reichart, D.E. & Mészáros, P. 1997, *ApJ*, 483, 597
 Schaefer, B.E., Deng, M. & Band, D.L. 2001, *ApJ*, 563, L123
 Schmidt, M. 2001, *ApJ*, 552, 36